SEEDS

Levels of Service / Cost Estimation Study

Working Paper Two:

Cost Estimation by Analogy Model

May 15, 2002

G. Hunolt, SGT, Inc.

Outline

1 Introduction

This working paper is the second of a set of papers that describes the SEEDS (Strategic Evolution of Earth Science Enterprise Data Systems) Levels of Service (LOS) and Cost Estimation (LOS/CE) study. The study goal is to develop a cost estimation model and coupled requirements and levels of services to support the SEEDS Formulation team in estimating the life cycle costs of future Earth Science Enterprise (ESE) data service providers and supporting systems, where 'data service provider' is used as a generic term for any data/information related activity. The set of working papers is intended to serve as a vehicle for coordinating work on the project, obtaining feedback and guidance from ESDIS (Earth Science Data and Information System project) SOO (Science Operations Office) and the user community, and as embryos of reports that will be produced as the task proceeds.

As working papers, each version of each working paper that appears represents a snapshot in time, with the work in various stages of completion; readers should expect loose ends and inconsistencies especially in the early stages of the project. As work progresses the content (and sometimes the organization) of the working papers will change reflecting progress made, responses to feedback and guidance received, etc.

1.1 Introduction to Working Paper 2 - Cost Estimation by Analogy Model

This second working paper of the set describes the Cost Estimation Tool and the underlying cost estimation by analogy model that is being developed by SGT for this study. (The Cost Estimation Tool is simply the packaging of the model in a usable form, including a user interface and report generating capability.) This paper will evolve extensively as the work progresses. Its initial focus will be on a conceptual description of the tool and the model, and how the model and the cost estimating relationships it will use are expected to develop, requirements and operations concepts including scenarios showing how the tool will be used, goals and plans for the demonstration prototype. A major update to this working paper will be provided by June 30, 2002 that will address plans for development beyond the demonstration prototype and for test and evaluation of the prototypes, including "tire-kicking" by users.

As a part of the LOS/CE Study, and in parallel with the effort described in this working paper, SGT is examining COTS cost estimating tools (e.g. parametric cost models) to see if one or more of these might be better for certain aspects of costing than the cost model to be developed during the study, or valuable for use in producing alternative cost estimates for some or all aspects of costing. At least some COTS tools can be integrated with other software such as Excel, and so it may be possible to deliver a cost estimation tool with an integrated COTS component. In any case, the most practical approach will be taken to facilitating the use of any selected COTS tool in conjunction with the model developed by the study.

As the needs of the ESE science and applications program evolve, and hence the ESE roles and missions for data service providers evolve, and as information technology that touches all aspects of every data service provider and the user community evolves (e.g. the Grid distributed computing approach), the data service provider reference model and the cost model will evolve. The content of this paper can only represent a snapshot in time - and indeed a snapshot that is in part tied to current and recent past experience with working data service providers. If the cost estimation tool (and the underling data service provider model) proves useful, it will have to be maintained and revised perhaps dramatically to preserve or improve its usefulness over time.

Technological changes and the evolving requirements of science and applications users, perhaps especially in the access and distribution area, call into question the ability of the cost estimation tool to make reliable cost estimates for the future based on current and recent past experience. The problem is acknowledged, and as the tool is developed and tested and its sources of error are analyzed, this aspect will not be ignored.

In addition to evolving with changing ESE program needs, the cost estimation by analogy model (and the data service provider model) will be improved in successive iterations as the comparables database grows and includes more new activities, and with lessons learned derived from use of earlier versions of the model.

Section 2 discusses the objective, scenarios, and requirements for the Cost Estimation Tool. Section 3 describes the output to be provided by the Cost Estimation Tool. Section 4 discusses development of the cost estimation by analogy model that is the heart of the Cost Estimation Tool, and discusses the first phase of cost estimation tool development, the demonstration prototype.

1.2 Definition of Key Terms

This section defines a few key terms that will be used frequently in the remainder of this working paper.

As noted above, the term 'data service provider' is used as a generic term for any data/information related activity, such as a data center (e.g. the EOSDIS Distributed Active Archive Centers (DAACs), a flight project data system (e.g. the MODIS Adaptive Processing System (MODAPS) or TRMM Science Data and Information System (TSDIS). A data service provider provides all or some subset of the functions described by the general data service provider reference model (see Working Paper 3, "Data Service Provider Reference Model - Functional Areas"), which include data ingest, processing, archive, distribution, etc. The scale and scope at which these services are provided depends on the particular ESE role or mission responsibility of the data service provider.

A data service provider is an element of, or is operated or hosted by, an ESE or ESE funded organization, which might be dedicated to supporting a single data service provider, a number of data service providers, or a combination of a data service provider and other different activities.

The term 'data service provider' is also used in an abstract sense in the context of logical data service provider types, as described in Working Paper 6, "ESE Logical Data Service Provider Types", which are essentially functional groupings, subsets of the general data service provider reference model, organized around an ESE general role or mission. Logical data service provider types do not necessarily map one to one to physical entities, i.e. to actual operating data service providers. A real world physical data service provider, such as a DAAC, often will embody more than one of the logical data service provider types, depending on the complexity of its specific ESE role or mission.

The term 'data service provider activity' refers to the work performed by a data service provider to meet the needs of a particular project or research or applications effort, regardless of whether the data services provider is an integral part of the project (such as a flight project data system like TSDIS or MODAPS) or whether the activity is performed by an organizationally or physically separate data services provider, or a hybrid split between those two cases (such as when a flight project does its own product generation but 'subcontracts' archive and distribution to a DAAC). A working data services provider may engage in a single activity, or multiple activities if it supports multiple flight projects or research / applications efforts, perhaps in addition to a core ESE data management mission.

In the most complicated case, a single ESE organizational element or ESE funded organization may serve as an ESE data service provider, embodying a number of logical data service provider types, and within each type, perform a number of activities supporting a number of projects or research / application efforts.

2 Cost Estimation Tool Scenarios and Requirements

This section discusses use case scenarios and requirements for the Cost Estimation Tool, where the term 'Cost Estimation Tool' is used to mean the cost estimation by analogy model packaged in a useable form, i.e. provided in a package that can be started up, can receive a set of inputs, run, and produce a set of outputs. An objective of the study is to provide the tool in as readily useable form as possible, for example as an Excel spreadsheet workbook that could loaded and used on any PC or Macintosh platform equipped with Excel.

As noted above, the Cost Estimation Tool is needed to enable the SEEDS Formulation Team to estimate the cost impact for various architecture trades, and to provide NASA Headquarters with estimates of the costs for implementing varying ESE mission profiles and implementation options. The Formulation Team also requires that the tool be packaged so that it can be provided to ESE scientists for their use in estimating the costs for developing and operating the science data ground system for their proposed mission.

The remainder of this section examines the particulars of this objective and what the Cost Estimation Tool must be able to do to meet it. Section 2.1 presents the use case scenarios, and Section 2.2 presents general requirements for the Cost Estimation Tool.

2.1 Cost Estimation Tool Scenarios

This section describes scenarios that describe how the Cost Estimation Tool would be used.

Several categories of use are envisioned:

1) The first is use by a flight project or science team to estimate the costs of implementing and operating a set of data management functions required to support their project or research effort. (This would constitute a single data service provider activity.)

2) The second is by an existing data center that has been asked to estimate the costs of adding an additional set of data management functions, perhaps to meet the needs of a new flight project or research effort. (In this case the data center would be adding an additional data service provider activity to those which it already performs.)

3) The third is by an ESE program office wishing to make overall estimates of implementation and operating costs for a constellation of ESE data service providers operated by a number of ESE or ESE funded organizations, collectively performing all of the data service provider activities required by the ESE program (i.e., supporting all ESE flight projects, research / applications efforts, and general data management needs). The ESE program office may wish to examine 'architecture trades' - alternative mappings of functions and mission responsibilities to ESE organizations, perhaps to make long term budget estimates.

Each of these categories of use will be discussed below.

2.1.1 New Project or Research / Applications Effort

Assume that a new flight project or research effort is being proposed in response to a NASA Announcement of Opportunity (AO) or other solicitation vehicle. The group developing the proposal examines its need for data management support - i.e. decides what functions it requires and what service needs it has for each (i.e. what levels of service it needs). It also puts together a description of its mission requirements for data management support - e.g. best quantitative estimates of what data will be received, produced, distributed (details sensor and ancillary data streams, products to be generated, distribution to team members or other users) etc.

The group is now ready to use the Cost Estimation Tool, and it is assumed that one group member proceeds as follows:

a. The user activates the tool, using a stand-alone distributed version or a web accessible version.

b. The user selects from a menu of functions those that are needed to meet the project's needs, and, for each function:

1) The user selects levels of service to be provided, as applicable or needed for the particular function.

2) The user provides quantitative mission detail as applicable for the particular function.

3) The user provides re-use factors, to account re-use of existing capabilities, if any.

c. The user provides control parameters such as implementation start date, operations start date, projected activity end date, etc.

d. The user provides costing parameters such as inflation rate, labor rates.

e. The user runs the model to produce the life cycle cost estimate for the data service provider activity.

The cost estimate will be a life cycle cost including year by year development and sustaining engineering costs and operations staffing and costs projected over a number of years (see Section 3 for a description of the output of the cost model).

This scenario makes no use of the predefined set of logical data service provider types, allowing the user complete flexibility in selecting needed data service functions. But the user would have the option of judging that a particular type of data service provider was a good fit, and at step b in the scenario bringing up a template for that type, and then providing the information as 'filling in the blanks' in the data service provider template.

2.1.2 ESE Data Center Takes on a New Task

Assume that an ESE data center, already engaged in a number of data management activities, perhaps supporting one or more flight projects or research efforts, wishes to propose to perform an additional data management task in response to a NASA AO, or has been asked by another group preparing a response to a NASA AO for (for example) a flight project to propose to provide data management support. Guided by the functions that would be required of it (in the examples given, either by the AO or the group preparing the flight project response), the data center would assemble the description of the new task requirements for data management support - e.g. best quantitative estimates of what data will be received, produced, distributed (details sensor and ancillary data streams, products to be generated, distribution to team members or other users), etc. The data center will also determine what ability it will have to "reuse" its existing infrastructure to support the new task (staff, systems, facility, etc.).

The data center is now ready to use the Cost Estimation Tool, and it is assumed that one data center staff member proceeds as follows:

a. The user activates the tool, using a stand-alone distributed version or a web accessible version.

b. The user selects from a menu of functions those that are required to perform the new task, and, for each function:

1) The user selects levels of service to be provided, as applicable or needed for the particular function.

2) The user provides quantitative mission detail as applicable for the particular function.

3) The user provides re-use factors, to account the data center's re-use of existing capabilities, if any.

c. The user provides control parameters such as implementation start date, operations start date, projected activity end date, etc., for the task.

d. The user provides costing parameters such as inflation rate, labor rates.

e. The user runs the model to produce the life cycle cost estimate for the data service provider activity.

The cost estimate will be a life cycle cost including year by year development and sustaining engineering costs and operations staffing and costs projected over a number of years (see Section 3 for a description of the output of the cost model).

This scenario makes no use of the predefined set of logical data service provider types, allowing the data center complete flexibility in selecting needed data service functions according to the requirements given it by the AO or the flight project (for the two possible cases noted). But the user would have the option of judging that the data center was a good fit for a particular type of data service provider, and at step b of the scenario bringing up a template for that type, and then providing the information as 'filling in the blanks' in the data service provider template, in that way 'tuning' the template for the data services provider type to meet the given requirements.

2.1.3 ESE Architecture Trade Study

This section describes the use of the Cost Estimation Tool to support ESE architecture trade studies.

In this context, the term 'ESE architecture' means a collection of physical entities, i.e. ESE or ESE-funded organizations, performing a set of data service provider activities that in their aggregate meet ESE's requirements for data management and services support to its flight projects and research and applications programs.

An architecture would be developed by analyzing the complete set of ESE program requirements for missions to be supported, science and applications efforts to be supported, data holdings to be maintained, the data service needs of various user communities within and without the ESE program, etc., and determining a set of ESE data service providers (and their required interconnections) able to meet the various ESE mission requirements. There will be many possible architectures, i.e. many possible configurations of ESE data service providers, that will meet a given set of ESE mission requirements, and hence the need for the ability to analyze trades between them to select the architecture to implement.

ESE mission requirements and hence the needed set of ESE data service providers will vary over time, given factors such as the launch dates for flight missions, the phasing of efforts in the science program, possibly the rotation of data into non-NASA long term archives, etc. As a result there will not only be many possible architectures at any given time, but also many possible paths for the evolution of the overall ESE architecture as time goes on. That is to say there will be many, time varying configurations of ESE data service providers that comprise possible ESE architectures. These may represent different approaches to consolidation of new mission requirements at ESE data service providers, different assignments of level of service requirements, different consolidations of new mission requirements with ongoing (i.e. EOSDIS) mission requirements, etc., and combinations of the above.

For each possible architecture, the cost estimation tool would generate cost estimates for the individual components and allow a roll-up of these to accumulate to an overall estimated cost for the candidate architecture. This would then be one factor taken into account in consideration of the trades between possible architectures (along with complexity, robustness, etc.).

Assume that an ESE data services architect wishes to generate a cost estimate for one of the many possible architectures. The first step is for the architecture to be defined as a mapping of data service provider activities to existing or new (presumably placeholders for winners of future competitions) ESE organizations. To simplify this task, the architect can use the defined logical data service provider types, and think of the problem as assigning data service provider activities or mission responsibilities (e.g. for supporting a flight project, supporting a research effort, providing general access to a large data collection, etc.) to a set of instances of data service provider types, and then mapping the instances of data service provider types to existing and/or new ESE organizations.

The ESE data services architect must also have specific mission requirements for the various flight projects, science efforts, etc., that comprise the ESE program. While an individual project (as in the cases described above) could be expected to know / forecast its own requirements in some detail, the ESE data services architect is likely to have to make rough estimates. This would reduce the accuracy of the result for each individual component of the architecture, and the error of estimate of the ESE architecture as a whole would be greater than that associated with an individual estimate produced by a flight project or data center as described in the previous sections.

Now the ESE program office user has a set of logical data service providers defined, and a set of organizations to which these are mapped, and mission requirements for flight projects, science / applications efforts, etc. The user could use the Cost Estimation Tool as follows:

a. The user could first use the Cost Estimation Tool to obtain an estimate for each logical data system provider as if it were to be physically a stand-alone entity performing a specific data service provider activity - i.e., mapped to one ESE organization dedicated solely to it, following the scenario in Section 2.1.1 above. The sum of all these would provide an aggregate ESE level cost estimate that would represent a worst case - having a separate physical ESE organization for each logical data service provider activity would result in no savings from reuse of infrastructure.

b. The user would then apply the mapping, i.e. the architecture:

1) For each ESE organization that as an element of the architecture, the user would select one of its logical data service provider functions as a base, and produce an estimate for it as described in Section 2.1.1.

2) Then the user would make separate estimates for the incremental costs of adding each of the other data service provider activities following the scenario described in section 2.1.2, until the total for that organization was reached. This would include taking re-use into account.

3) The user would repeat the process for all of the ESE organizations comprising the architecture.

c. The user would sum up the estimates for the individual organizations and arrive at an estimated cost for the specified architecture, including an overall error of estimate.

d. The user could then modify the architecture, and redo the process, to obtain another estimate for comparison.

2.2 Cost Estimation Tool Requirements

This section outlines the basic requirements the Cost Estimation Tool (a.k.a. the Cost Tool) shall be capable of meeting, following the scenarios in Section 2.1. The requirements are divided into two groups, functional and implementation.

2.2.1 Functional Requirements

The Cost Tool shall be based on a cost estimation by analogy approach, supplemented by parametric techniques as needed or advantageous.

The Cost Tool shall estimate full life cycle costs for a data service provider activity, from the beginning of implementation through a specified operational life time.

The Cost Tool shall estimate costs for all areas of cost, including implementation, operations, maintenance and management staff, COTS hardware and software, custom software, logistics and supplies, etc. See Section 3 for a detailed description of the required Cost Tool output.

The Cost Tool shall provide errors of estimate for each projected cost.

The Cost Tool shall allow the user to select functions to be included in the estimate for a data service provider activity and provide parameters for each function appropriate to the data service provider activity's mission or ESE role.

The Cost Tool shall, as an alternative mode of operation, allow a user to select from a set of logical data service provider types, each including a pre-selected set of functions.

The Cost Tool shall support the estimation of costs for a data service provider activity as either a new, stand-alone activity (such as a new data system to be implemented as part of a NASA flight project) or as an incremental increase to an existing data service provider activity (such as the incremental costs to an existing NASA data center to provide archive and distribution support to a NASA flight project), in which case the Cost Tool user shall provide appropriate re-use factors.

The Cost Tool shall support subcontracting or partnering, such as allowing for a new NASA flight project arranging for an existing NASA data center to provide archive and distribution support.

The Cost Tool shall support generation of cost estimates for an ESE data services architecture, including overall error of estimate.

2.2.2 Implementation Requirements

This section describes implementation requirements for the Cost Estimation Tool.

The Cost Estimation Tool shall be capable of stand-alone operation on any reasonably sized PC or MacIntosh platform, as an Excel workbook application or equivalent.

The Cost Estimation Tool shall be fully documented, including a users' guide.

The precise form of the cost estimation tool is TBD as of now, but might include an integrated COTS tool, or the package might include both an Excel based tool and a separate COTS tool or tools. In any case, the package will be documented as a whole, with a users' guide covering the entire package, supplemented by COTS documentation as needed.

3 ESE Data Service Provider Cost Estimate Content

This section describes the content of cost estimates to be provided by the Cost Estimation Tool, i.e. a statement of what the cost estimate provided by the tool will consist of. This description of the model output will drive requirements for the combination of input and computation needed to produce it.

The Cost Estimation Tool must provide estimates of implementation and operating costs, over a specified lifetime, for a data service provider activity (as a stand-alone or increment). The implementation period costs must include hardware purchase, custom software development and COTS purchase, integration and test costs, and facility preparation costs. The operating period costs of a data service provider activity must include hardware maintenance, continuing COTS support, sustaining engineering, operations, recurring facility costs, supplies such as storage and distribution media, and must allow for the possibility of 'technology refresh'. Implementation and operations period staff costs must allow for reasonable management staffing, and labor rates must allow for overhead and inflation adjustments.

Tables 1, 2, and 3 below are together an initial example of what the cost estimation output might look like. The categories would be defined in detail below. Note that the actual number of years for which costs would be estimated (shown as seven in the example) would be selectable as appropriate for actual cases.

Table 1 - Draft Sample of Cost Estimation Output - Implementation Period Costs

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Total
Estimated Implementation Costs								
Management Staff, FTE								
Management Staff Cost								
Development Staff, FTE								
Development Staff Cost								
Hardware Purchase								
COTS Software Purchase / License								
Facility Preparation								
Total Implementation FTE								
Total Implementation Cost								

The cost estimate example shown above and below contains some FTE lines that would be generated by the model in the process of producing the cost estimate. Other such lines could be added, such as SLOC to be developed and maintained.

There are a variety of "workload" parameters that could be presented in conjunction with the cost estimate. These could include those that characterize the mission of the ESE data service provider, which would have been provided as input to the cost estimation, such as flight mission to be supported, input data streams, output product streams, etc., appropriate to the type and particular mission of the data service provider.

Table 2 - Draft Sample of Cost Estimation Output - Operations Costs

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Total
Estimated Operations Costs								
Management Staff FTE								
Management Staff Cost								
Technical Coordination Staff FTE								
Technical Coordination Staff Cost								
Sustaining Engineering FTE								
Sustaining Engineering Cost								
Engineering Support FTE								
Engineering Support Cost								
Operations Staff FTE								
Operations Staff Cost								
Development FTE								
Development Staff Cost								
Recurring Network / Comm Cost								
Recurring COTS S/W Cost								
Hardware Purchase Cost								
Hardware Maintenance Cost								
Supplies Cost								
Recurring Facility Cost								
Total Operations FTE								
Total Operations Cost								

Table 3 - Draft Sample of Cost Estimation Output - Total FTE and Costs

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Total
Estimated Total FTE								
Estimated Total Cost								

All of the parameters shown in the tables above are defined in Working Paper 4 “General DSP Reference Model - Model Parameters”, as are also the required input parameters and all of the internal parameters used by the cost estimation by analogy model.

4 Development of the Cost Estimation by Analogy Model

This section discusses the development of the cost estimation by analogy model that is the heart of the Cost Estimation Tool.

4.1 Cost Estimation By Analogy

The cost estimation by analogy technique is based on the idea that reasonably reliable estimates for the cost of a future data activity (either by a new organization formed for that purpose or as an increment to the data activities of an existing organization, or some combination thereof) can be based on an analysis of the past history and experience with other similar data activities.

Contained in this simple statement are some assumptions that must be taken fully into account:

1) That a sufficient sample of reasonably applicable cases exists on which to base an estimate.

2) That cases are either applicable in implementation and operation approach as well as function and workload, so that the effort required for the cases can be taken as suggestive of the effort that will be required for a new case to be estimated.

The first assumption is important when statistical techniques such as regression are considered; if there is too small a sample the results will be unreliable or entirely useless, as will be indicated by the probable errors of estimate that will accompany the estimates, and by the results of tests on independent cases.

The second assumption reflects the concern that a project that might be nearly identical in terms of the nature of the data activity (function and workload) to be estimated but might have been done (implemented and/or operated) by an approach so different as to compromise partly or completely its value as a data point for producing an estimate for a new activity. Attention must be paid to trends that could follow changes in approach that might provide a basis for an extrapolation into the future.

An important point that must be made at the outset is that the model will not be estimating future costs on the basis of past costs. It is indeed almost a misnomer to call the model a 'cost model' because the real basis for comparison with cases is staff effort and system capabilities. Year by year costs are only added as a final step. A year by year effort estimate is first produced, and then priced out by application of labor rates and inflation. Similarly projections of required system capabilities are made, and then priced out through use of system capability vs projected cost curves. Other non-staff elements of cost are handled in like manner. Finally all factors are summed to produce the final output, the year by year life cycle cost estimate.

4.2 General Development Considerations

This section outlines some general considerations regarding the approach being taken to developing the cost estimation by analogy model.

The model will evolve over the life of the project. This evolution will be driven by a number of forces: 1) the building of the comparables database, 2) the feedback obtained from users evaluating prototypes of the Cost Estimating Tool, 3) the feedback gained from experience with actual use of the Cost Estimating Tool, and 4) the progressive growth of the comparables database, as new cases are added and as the information about existing cases is updated.

The building of the comparables database will drive the evolution of the model. This is because the cost estimating relationships used by the model are dependent on, or constrained by, the state of the comparables database. The types of cost estimating relationships to be used by the model are described in the next section. Those relationships which draw on the comparables database - the true estimation by analogy relationships - have to be developed through analysis of the available data. The state of the available data will develop slowly as the information collection process goes on - i.e. as the comparables database is gradually built. In the case of some parameters, a sufficient number of comparable cases will be accumulated to enable statistical relationships to be used. For other parameters this will not be the case, and either reasonable arithmetic approximations will be used or the parameters will have to be dropped. Thus the model has to be flexible to accommodate changes to the cost estimating relationships as more is learned about the data that will be available and various possible combinations of parameters are tested to see which combinations yield the strongest relationships.

At first only simple relationships will be employed, but as development proceeds the use of non-linear relationships will be explored, and perhaps tools / techniques that evaluate the relative 'distance' of the input case to the members of the set of comparables to produce a better estimate.

The general approach to the development will be to work top down, to try to come up first with the simplest possible set of cost estimating relationships, even dummy placeholders, but a working model that demonstrates how the model will run, the user interface, the output to be produced.

If the model is thought of as a Fortran program - the development approach is to have a working main program with input and output subroutines routines and dummy computational subroutines that runs so users can see and provide feedback on how it works. Imagine that each arithmetical relationship or comparables-based CER is a subroutine. The model will begin with an initial set of subroutines, each with an output and set of inputs, that uses a very simple method of doing its computation. Then, as time goes on, each subroutine will be replaced with more advanced versions which produce the same outputs but use better computational approaches, maybe a linear regression type relationship that someday might itself be superseded by a different one based on a better set of inputs or a non-linear relationship. The intent is to have a "working" model right from the start, that will gradually get better as its parts are improved.

4.3 Cost Estimating Relationships and Model Parameters

The output of the Cost Estimation by Analogy Model (or the tool which embodies it) is a list of parameters (see Section 3 above). Each of these output parameters is related to (computed from) combinations of other parameters in several steps, ultimately tracing back to either input parameters provided by the model's user or to parameters obtained from the comparables database, or combinations of the two. The term 'chain' will be used for the full sequence from an output parameter back to its ultimate inputs. Each chain consists of one or more 'links'; i.e. each chain from an output parameter back to the user input and/or comparables input from which it is ultimately derived can be thought of as one or more steps, each a link in the chain. Each link consists of an [output][process][input] sequence, with the links being connected by an overlaps between inputs and outputs. For example if output parameter A was computed from intermediate parameter B which in turn was computed from an input parameter C, then the chain would have two links, [A][process][B] and [B][process][C], with the connection between them being the parameter B which is the input to the second step and the output from the first step.

The [process] portion of each link contains a rule for computing the output from the input, known generically as a "cost estimating relationship" regardless of whether or not the output is a cost per se. The model employs three kinds of cost estimating relationships (CERs) - 'plug value', 'arithmetic', and 'comparables based'. Each of these kinds is described below.

Detailed documentation of the chains and links showing the relationships between the parameters and describing the computational steps will evolve as the project progresses. Copies are available by request.

4.3.1 CERs of the First Kind - 'Plug Values'

The first kind of CER is the 'plug value'. Plug values are constants plugged into the value when there is no better way of computing the output parameter. For example, a parameter may be defined to capture the level of effort required for participation in the process of developing and maintaining standards. A level of 0.25 FTE might be assessed as reasonable for this, in the absence of good documentation of actual levels of effort in past situations, or any basis for computing a level. That value, 0.25 FTE, becomes a plug value for the parameter.

4.3.2 CERs of the Second Kind - 'Arithmetic'

The second kind of CER is 'arithmetic'. In this case there is a simple arithmetic relationship between the output and its input(s). For example, suppose that an output parameter is the cost for management staff, and the inputs are effort in FTE, management labor rate, and the inflation rate. The arithmetic process to get the cost would be to multiply the FTE by the labor rate and apply the inflation adjustment to the result.

4.3.3 CERs of the Third Kind - 'Comparables Based'

The third kind of CER is 'statistical'. The output parameter is computed by a relationship that involves the data (i.e. one or more parameters) from the comparables database. The relationship may be based on linear regression, a non-linear relationship, or some more complex technique. An error of estimate will accompany the result.

The information in the comparables database (though assembled on a site by site basis) will be used on a parameter by parameter basis within the reference model's functional areas. The 'best fits' for a projected new data activity's ingest area might includes cases that were not good fits for other areas, etc.

4.4 Demonstration Prototype Cost Model - Development Approach

This section described the approach being taken to development of the demonstration prototype of the Cost Estimation Tool and the cost estimation by analogy model it contains.

4.4.1 Objectives for the Demonstration Prototype

These are the minimum objectives for the demonstration prototype.

1. The demonstration prototype will show how the Cost Estimation Tool will work, how a user will use it, how the scenarios in Section 2 above will be realized. The prototype will show a user picking from a general function list - i.e. be based on the general reference model and not show the data service provider subsets. The model has to do a complete execution, regardless of what simplifications are necessary at this point.

2. The demonstration prototype will use a partial, very limited comparables database based on the benchmark study data plus whatever else can be collected from about a half dozen sites.

4.4.2 Not Objectives for the Demonstration Prototype

These are things the first, demonstration prototype will not be capable of, presented in order to be clear about expectations.

1. The demonstration prototype will not produce useful results. The ability to produce useful results depends on the database of comparables being as large as possible, allowing the best CERs, and in the prototype timeframe the information collection and building of the comparables database will have just begun. Results will not be tested against independent cases - that will come later when more data is collected and some cases can be held aside for such testing.

2. The demonstration prototype will not show how the model can be used to estimate the costs of various possible SEEDS architectures - combinations of sites that are each combinations of one or more data service providers performing data activities.

3. The demonstration prototype will not show 'subcontracting' or 'teaming' - that will come later.

4.4.3 Steps to Develop the Demonstration Prototype

1. Develop an initial version of the chains and their links, the relationships between the output, internal, and input parameters, with placeholders for the CERs that will be needed.

2. Develop the schema for the comparables database and a site 'template' (or 'fill in the blanks' form), populate from the benchmark data and site data as possible (starts and goes on in parallel for at least a year).

3. Create a User Interface description, input and output, i.e. what the user will see. Reference the scenarios in Section 2 above.

4. Develop an initial set of CERs.

As far as the inputs go, as the comparables database is built it will be seen which of the possible inputs shown in the parameter matrix can actually be collected or collected in sufficient number to be used. The demonstration prototype will have to make do with whatever is available.

As far as the 'comparables based' computation by the prototype goes, given the limited data set that will be available, it is expected that only linear relationships will be used, e.g. a simple linear regression technique with error of estimate, an Excel function. Thus the demonstration prototype will have a set of linear equations for comparables-based CERs. There will doubtless be a number, perhaps a large number of cases where there will not be sufficient data to use even simple regression. In these cases, as placeholders, arithmetic relationships based on a (documented) assumption or two.

Other CERs will be 'plug value' or 'arithmetic'.

4.5 Development Beyond the First Prototype

This section will be developed between the first version of the working paper and June 30, 2002. It will address the working prototype and operational versions of the Cost Estimation Tool. The basic approach will be a progressive refinement of the model and its CERs as the comparables database grows and as results of testing are taken into account. "Tirekicking" of prototypes of the Cost Estimation Tool by users will be undertaken and is expected to provide valuable feedback.

References and Acronyms

The References Section and the Acronym List for all of these Working Papers is in the document

"References and Acronyms for the Levels of Service / Cost Estimation Working Papers ".